Is ‘Cyberwar’ Fought with Weapons?

Lisa M. Cohen 2021-09-09T08:00:16

How do – and should – we classify offensive cyber capabilities? Particularly in the context of international humanitarian law, the question whether offensive cyber capabilities are weapons, means of warfare, or methods of warfare is not just semantics but has legal implications. Yet, there is great terminological inconsistency.

The most prominent term is arguably that of ‘cyber weapons’. It is used in the Tallinn Manual 2.0 (a non-binding comprehensive academic study on the application of international law to cyber operations), and military manuals (e.g. Denmark), by states (Brazil, Egypt), the International Committee of the Red Cross (ICRC), tech companies, government officials, the press, and academics. In contrast, Australia and the US largely refrain from calling offensive cyber capabilities ‘weapons’ (see e.g. the Air Force Instructions, treating weapons and cyber capabilities as two distinct categories). At the same time, the term “cyber means and methods of warfare” can be found in many recently published cyber position papers (see e.g. Germany, and Finland).

Against the background of these terminological inconsistencies, this article will examine the legal relevance of and possible approaches to the classification of offensive cyber capabilities to then propose criteria according to which these capabilities could be classified.

The Terminology and Legal Implications of the Classification

Offensive cyber capabilities are resources, skills, and operational concepts used to manipulate, deny, disrupt, degrade or destroy targeted communications and information systems and achieve strategic, political, or military objectives in or through cyberspace. They include, but are not limited to, “computer code that is used, or designed to be used, with the aim of causing physical, functional, or mental harm to structures, systems, or people.”

While the term ‘cyber weapons’ is very much *en vogue*, there is no authoritative or globally acknowledged definition of ‘cyber weapons’ (see here, here, and here). The inflationary and often unreflective categorization of offensive cyber capabilities as ‘cyber weapons’ neglects the specifics of the capability in question and leads to confusion about the legal rules applicable thereto.

In the context of international humanitarian law (IHL), the terms ‘means of warfare’ and ‘methods of warfare’ are much more pertinent. Importantly, ‘means of warfare’ is commonly understood to encompass weapons and weapons systems (cf. ICRC Commentary to Additional Protocol I, Rule 103 of the Tallinn Manual 2.0, and Switzerland’s cyber position paper, p. 9). In this sense, whatever falls within the ambit of the above definition of ‘cyber weapons’ also falls under the ambit of ‘means of warfare’. The term ‘methods of warfare’, on the other hand, “designates the way

or manner in which the weapons are used" and "comprises any specific, tactical or strategic, ways of conducting hostilities that are not particularly related to weapons and that are intended to overwhelm and weaken the adversary" (see here, cf. Rule 103 Tallinn Manual 2.0).

While the review obligation and the precautionary principle under Articles 36 and 57(2)(a)(ii) Additional Protocol I, respectively, apply to both means (encompassing weapons) and methods of warfare, the classification of cyber capabilities as either means or methods of warfare makes a crucial difference regarding the law of neutrality. The latter forbids belligerents "to move […] convoys of either munitions of war or supplies across the territory of a neutral Power" (Article 2 Hague V) and obliges neutral states to prevent their territory from being used by the belligerents (Article 5 Hague V). The Tallinn Manual 2.0's Rule 151 states that "physically transporting cyber weapons [and] transmission of cyber weapons across cyber infrastructure located in the neutral State" falls under the prohibition of Article 2 Hague V.

Problematically, due to the interconnectedness of cyberspace, code used for offensive cyber operations will almost always be routed through neutral territory and civilian information and communications technology (ICT) infrastructure. For the same reason, its path and spread are nearly impossible to control once employed. Even the Stuxnet worm, employed against an Iranian nuclear facility, though considered carefully designed and precise, did not only affect the targeted system but spread to several unintended targets across multiple countries. During distributed denial-of-service (DDoS) operations, Botnets, *i.e.* networks of hijacked internet-connected devices which are remotely controlled and operated to perform a certain task, are employed to flood a target system with requests to overload and disrupt it. The massive number of requests likely also overwhelms and (temporarily) incapacitates the systems it is routed through. Accordingly, the employment of offensive cyber capabilities that classify as means of warfare would almost always violate the law of neutrality. This would render their employment virtually impossible.

On the other hand, if offensive cyber capabilities were seen as methods of warfare, the law of neutrality would not prohibit states from transmitting code used for offensive cyber operations through neutral territory: The transmission of code would be a permitted use of ICT infrastructure in the neutral state (cf. Article 8 Hague V). This would risk undermining the purpose of the law of neutrality – protecting neutral states and their nationals and preventing a further escalation of the conflict.

'One Size Fits All' versus Case-by-Case Approach

In accordance with the terminology chosen in the Tallinn Manual 2.0, most publicly available cyber position papers contain the notion 'cyber means and methods of warfare' (e.g. Australia, Germany, Finland, Australia, Switzerland, and the ICRC). While all documents lack an elaboration on criteria for distinguishing between means and methods, they indicate support for a classification on a case-by-case basis.

In contrast, a 'one size fits all' approach to the classification is now advocated by the Tallinn Manuals' general editor, Michael Schmitt. By comparison, he concludes

that offensive cyber capabilities lack a prevalent characteristic of acknowledged physical weapons – direct causation of the terminal effect on the target – because they "merely tr[y] to convince another computer to do something". Thus, all offensive cyber capabilities would constitute methods of warfare.

Considering the variety of types and technical means of cyber capabilities, a 'one size fits all' approach is overly simplified. Moreover, relying on characteristics decisive in the physical domain neglects the fact that cyber capabilities' nature, deployment, and ways of causing harm are fundamentally different from physical weapons. While the terminal effect (i.e. damage/destruction to objects, injury/death to persons) is a primary effect of physical weapons, it is usually a second-, or third-order effect of cyber capabilities. Primary effects of cyber capabilities include "the deletion, corruption, or alteration of data or the disruption of an adversary's computer network." The Stuxnet worm, for example, successfully took over the operation of centrifuges in an Iranian nuclear enrichment facility and caused the system to send faulty instructions (first-order effect), leading to the malfunctioning of the centrifuges (second-order effect) and ultimately the destruction of some centrifuges (third-order and terminal effect).

In the desirable event that loss of functionality is recognized as damage, the primary effect of Botnets used in DDoS operations, the (temporary) incapacitation of the target system with the amount of requests they send to the target system, and ransomware (like WannaCry) which encrypts targeted files rendering them unusable, would also be the terminal effect. The same is true for malware deleting data, if data is considered an object. Those capabilities have in common that – once launched – they operate and develop independently from human interaction, thus entailing an almost incalculable risk of spreading uncontrollably and harming unintended targets. One could refer to them as automated.

Different from these cyber capabilities are non-automated cyber capabilities, which are employed to gain access to and control over a system which can then be manually influenced (consider e.g. recent water plant incidents). Here, the terminal effect is not initiated by the primary effect of the offensive cyber capability. To cause harm, intermediary human interaction is needed to exploit the situation established by the primary effect.

Conclusion: A Proposal for Classification Criteria

Against this background, the capacity to inflict damage or destruction as a primary effect must not be an essential feature of a cyber weapon – because no authoritative weapons definition requires it, and it contradicts the mode of action of all cyber capabilities. The classification of offensive cyber capabilities should focus on technical means, the need of intermediary human interaction to cause harm, and on effects on the target system, the civilian population, and neutral states.

In line with this, offensive cyber capabilities would classify as means of warfare (weapons) if they 1) are automated, 2) have the capacity to initiate the process leading to the terminal effect on the target independently without human interference, and 3) are intended to and can cause a certain degree of harm (not

just inconvenience, irritation, or fear). Stuxnet, for example, would fall within this category.

Non-automated offensive cyber capabilities, on the other hand, where the situation created by virtue of the primary effect must be actively exploited (e.g. water plant incidents) to cause harm, would constitute methods of warfare. This also fits in with methods of warfare of the physical domain, like perfidy, where e.g. feigning surrender or improper use of distinctive emblems elicit confidence from the belligerent (primary effect) which is betrayed/exploited to cause harm.

While the criteria set out above will likely raise some intricate follow-up questions, they can be a starting point to discuss the classification of cyber capabilities independent from inadequate reliance on cyber capabilities' ability to demonstrate characteristics of physical weapons.

The "Bofaxe" series appears as part of a collaboration between the IFHV and Völkerrechtsblog.
